Memory System Design

This page presents a coherent view of the design of a static RAM (SRAM) memory system. It is intended to complement the presentation of this topic as it is covered on pages B-26 through B-31 of *Computer Organization & Design: The Hardware/Software Interface (Second Ed.)* by Patterson and Hennessy (Morgan-Kaufmann, 1998). This page is a complement to the presentation in the Patterson and Hennessy text, not a substitute for it.

The problem with the Patterson and Hennesy presentation is that it is inconsistent or incomplete as it crosses the three levels of the design: binary cells, SRAM ICs, and Memory System. The three levels are treated consistently in this document.

Binary Cell Design

We begin with the design of a circuit which can store one bit of a memory, a single binary cell. The text does not introduce this level separately, and it is not clear what the *Enable* inputs to the binary cells control when they are used in Figure B.23. The following diagram shows a clocked D latch augmented with a tristate buffer controlled by an *Enable* input to the latch:

### Clocked D–Latch with Output Enable

(One Binary Cell)
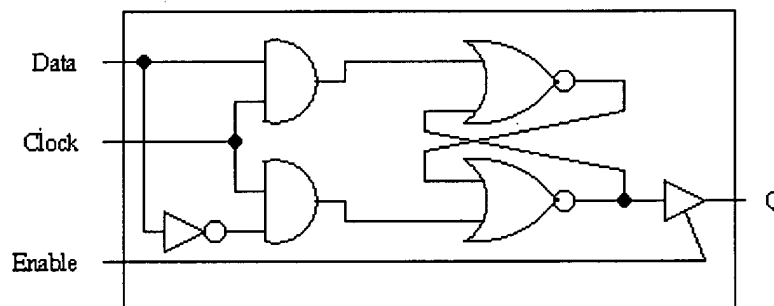
SRAM IC Design

Here is a version of Figure B.23 that is consistent with the block diagram given in Figure B.21 of the text. Each output of the decoder goes to two AND gates, one for reading and one for writing. In addition, the Chip Select input to the RAM IC is connected to the *Enable* input of the decoder, which is shown in the next figure. The other difference between this figure and B.23 is that it indicates how the word size of the IC would be expanded beyond two bits per word. Look at the right side of the diagram, and the labels for the data input and data output lines.
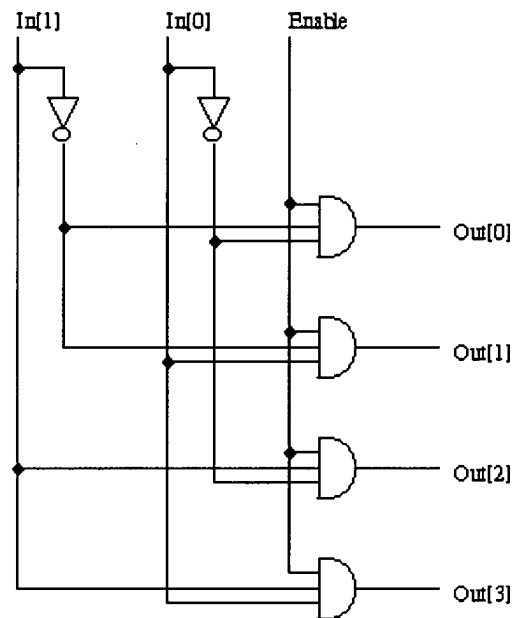
Four Word SRAM IC
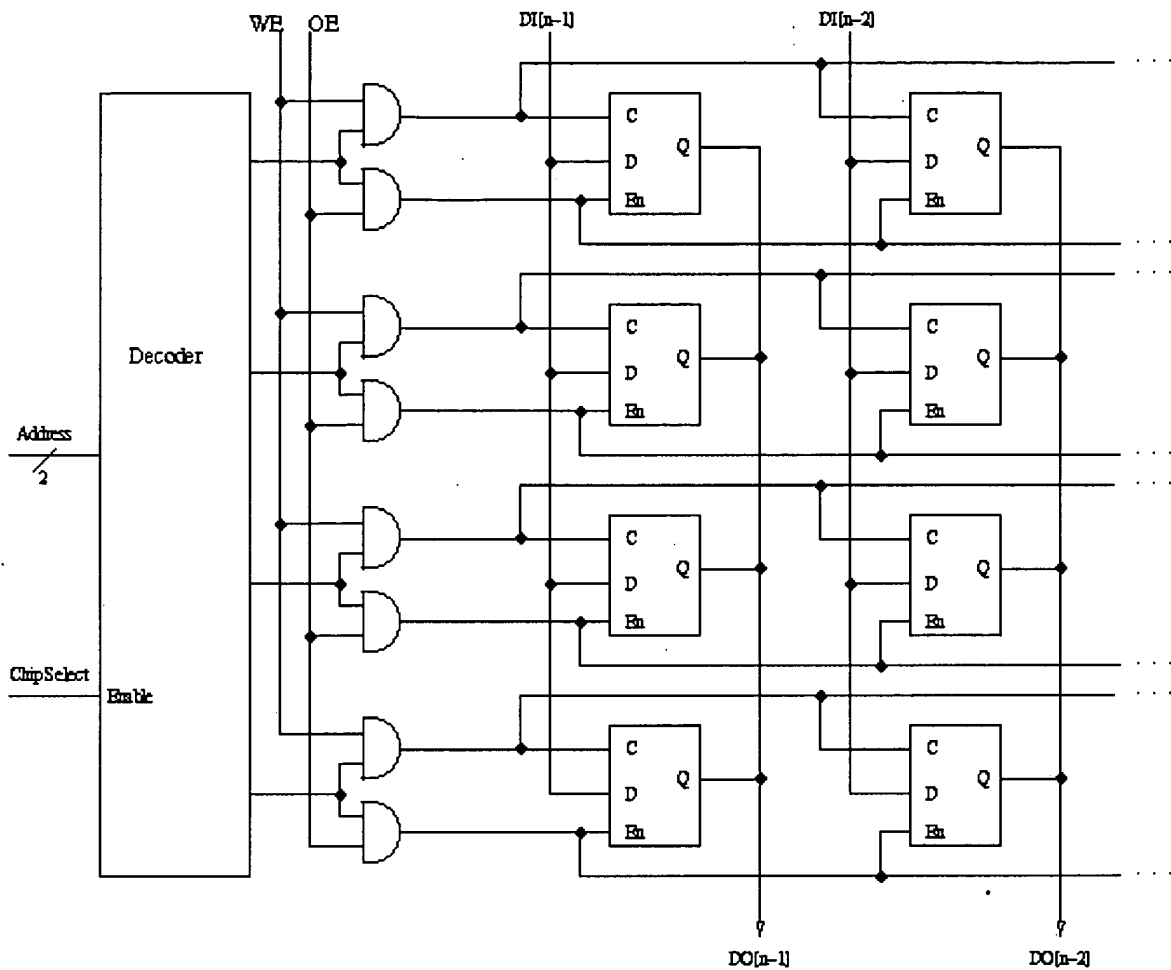
Structure of a 4x*n* SRAM.

Central to the design of the SRAM IC is that the entire chip can be disabled so that it neither writes or supplies output data. This is done by making the *Chip Select* input false. To accomplish this, we simply add an additional input to each of the AND gates inside the decoder part of the chip:

Decoder with Enable

Decoder used inside an SRAM IC.

Memory System

Figure B.24 of the text shows how an array of SRAM ICs can be connected to build a memory system. Unfortuantely, it is incomplete in that it does not indicate how write operations could be performed. Since the bits for several words (64) are stored at the same address in each IC, writing is complicated: to write into one word would require reading the locations of all the ICs (giving 64 words), the required word would then be changed, and all 64 words (including the one modified word) would have to be rewritten. This complexity is not actually a problem in a memory system that supports cache, since multiple words of primary memory are normally transferred between primary memory and cache memory anyway. It's just that Figure B.24 would be even more difficult to understand if the writing part of it was included.

As an alternative, here is the design of a 4Mx32 memory system built from 2Mx8 SRAM ICs. The upper part of the figure is a block diagram showing the number of address and data bits for the entire memory system.

Design a 4Mx32 memory system from 2Mx8 SRAM ICs
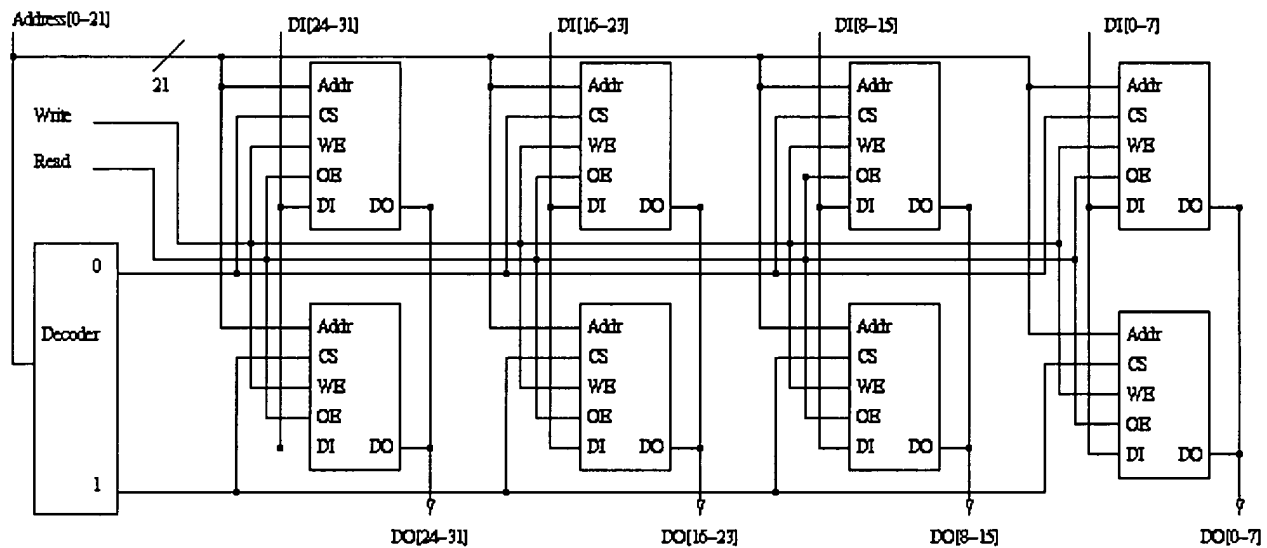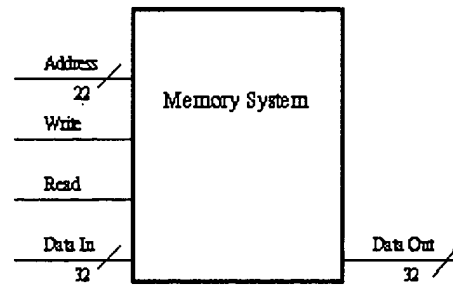
As this design illustrates, the dimensions of the matrix of SRAM ICs in the memory system can be derived as follows: The number of rows of ICs equals the ratio of the number of words in the memory system to the number of words per IC, and the number of columns of ICs equals the ratio of the number of bits per word in the memory system to the number of bits per word per IC. Of course, this structure applies only when both ratios are one or greater, which is not the case in Figure B.24 of the text. Also, note that the number of rows of ICs determines the structure of the decoder in the memory system, and the number of address bits that must go into this decoder.

Finally, we note that the designs presented here and in the textbook are intended only to show the principles of designing a memory system. In the real world, the situation is complicated by the need to provide a mechanism for implementing subsets of a memory system in a distributed fashion. In general, this means that the logic for decoding address bits into *chip select* wires will have to be partitioned across multiple memory modules or circuit boards.

Christopher Vickery
Queens College of CUNY